

Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/>

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication

Session 1aSCa: Distinguishing Between Science and Pseudoscience in Forensic Acoustics I

1aSCa3. Voice stress analyses: Science and pseudoscience

Francisco Lacerda*

***Corresponding author's address: Stockholm University, Stockholm, SE-106 91, Sweden, Sweden, frasse@ling.su.se**

Voice stress analyses could be relevant tools to detect deception in many forensic and security contexts. However, today's commercial voice-based lie-detectors are not supported by convincing scientific evidence. In addition to the scientific implausibility of their working principles, the experimental evidence invoked by the sellers is either anecdotal or drawn from methodologically flawed experiments. Nevertheless, criminal investigators, authorities and even some academics appear to be persuaded by the ungrounded claims of the aggressive propaganda from sellers of voice stress analysis gadgets, perhaps further enhanced by the portrays of "cutting-edge voice-analysis technology" in the entertainment industry. Clearly, because there is a serious threat to public justice and security if authorities adopt a naïve "open-minded" attitude towards sham lie-detection devices, this presentation will attempt to draw attention to plausibility and validity issues in connection with the claimed working principles of two commercial voice stress analyzers. The working principles will be discussed from a phonetics and speech analysis perspective and the processes that may lead naïve observers into interpreting as meaningful the spurious results generated by such commercial devices will be examined. Finally, the scope and limitations of using scientific phonetic analyses of voice to detect deception for forensic purposes will be discussed.

Published by the Acoustical Society of America through the American Institute of Physics

Voice stress analyses: Science and pseudoscience

INTRODUCTION

Scientific methodology is characterized by transparency and open self-criticism. It relies on hypotheses and models that can be proven wrong if they cannot provide appropriate answers when challenged on logical grounds. Science progresses by trial and error. The knowledge it generates is never definitive or private. Scientific knowledge is reviewed whenever it is proven not to provide logically correct answers. The strength of scientific methodology comes from its principles of open public character in combination with the demand of flawless logical argumentation. The fact that its proposed theoretical models are falsifiable but nevertheless resist public scrutiny strengthens the model, although there is in principle the possibility of being proven wrong in the future (Popper, 1992). The public character of the scientific endeavor, involving open presentation of data and an openness to possibility of having reached incorrect interpretations, is therefore in sharp contrast with the secrecy and “private-knowledge” character of pseudoscience. Unfortunately, the so-called voice stress analyses to be discussed in this session fall clearly in the category of pseudoscientific methods. Although the manufacturers attempt to give the impression that their methods rely on established scientific principles, they provide implausible explanations, refer to some sort of proprietary knowledge and their empirical evidence comes typically from flawed or irrelevant experiments. An important question is why these methods are nevertheless so popular among certain groups of criminal investigators. There are several possible answers to this question but their obvious common denominator is that those investigators do not have a basic understanding of the working principles of such voice stress analyses nor do they have the rudiments of basic experimental methodology. Another possible answer is that they do not have any ethic concerns provided the gadgets lead to “case solved” by whatever reason.

The following sections start by discussing generic aspects of physical forensic evidence, including acoustic signals. Focus is then shifted towards acoustic signals and the particular case of speech recordings, which are briefly discussed from the general perspective of providing useful information for speaker identification, in addition to direct linguistic content. Finally two working principles used in commercial devices – “voice stress analysis” (VSA) and “layered voice analysis” (LVA) – are briefly described and examined from a methodological perspective.

GENERIC FORENSIC EVIDENCE

Scientific Forensic Methodology

Reconstructing events and the timeline of their occurrence is one of the primary concerns of forensic investigations. In addition to the technical crime-scene analyses, the suspects’ accounts of the events as well as the witnesses’ observations and reports are essential contributions to build a grounded theory of what happened, when and, hopefully, why. The scientific analysis of physical items and the thorough objective study of available information are crucial components of a professional investigation that can stand the test of a fair public trial. Indeed, in a correctly conducted trial, it is the quality of the available information that ultimately limits the validity of the verdict. However, the criminal investigators’ path towards focusing on relevant information is not obvious from the beginning and must be guided by a broad, open-minded attitude, in dramatic contrast with the “experienced investigator’s hunch” typically portrayed by the film industry. In fact, successful investigations require freedom from scope-narrowing prejudices along with gathering of uncontaminated information and the systematic documentation of all steps taken. In other words, a correctly carried out criminal investigation must follow the same methodological principles that are used in scientific research to generate and prove theories emerging from systematic and well-defined observations. In both cases, the methodological stringency underlying error control and a step-by-step logical flow is essential to achieve a theory or explanation of facts that can stand the challenges of public scrutiny.

Forensic evidence is typically concerned with laboratory analysis of materials, documents or biological tissues that can be characterized in different ways. An essential feature of this type of analysis is that the evidence consists

F. Lacerda

of relatively stable materials that can be objectively measured and associated with certain components of the investigation. This is for instance the case of fingerprints or DNA analyses that can be used to establish who was present at the crime scene. Video or still images can also be analyzed and used to demonstrate that there is a strong link between a suspect and the perpetrator. Recordings of sound are yet another example of analyzable physical events that can help establishing a correct reconstruction of the events. However, usefulness of such recordings is critically dependent on the type of sound of interest and the purpose of the analyses. Speech recordings from the scene are obviously particularly interesting but they raise very specific and difficult analysis problems that the other types of physical information do not have. In contrast with fingerprints, DNA and images that reflect inherent stable anatomic and physiological individual properties, speech sounds are the result of human behavior and therefore much more loosely determined by the speaker's underlying anatomic and physiological properties. Thus, whereas the analysis of acoustic phenomena may provide some useful information about the physical dimensions and other properties of a crime scene the same is not necessarily true when only speech recordings are available. The reason for this is that most speech characteristics can, in principle, be intentionally altered by the speaker to the point of presenting surface acoustic properties that may be taken as coming from another individual. If the purpose of an acoustic analysis is to match acoustic samples of the suspect's voice with speech samples known to come from the perpetrator, it is necessary to demonstrate that the match is based on the acoustic signatures of the speakers' anatomy and physiology rather than on the part of the speech signal that might have been altered by the speaker. At first sight, this separation between inherent acoustic speaker characteristics and stylistic behavioral features might be theoretically possible. However, even such inherent acoustic characteristics are critically affected by subtle factors like muscular tension and compensatory articulatory movements which inevitably complicate the separation problem. In summary, speaker identification by instrumental analyses of speech samples is quite problematic because human speech is the outcome of a learned behavior that explores the respiratory (and part of the digestive) system's acoustic affordances, from which it is difficult to assess the system's underlying geometric characteristics that would be ultimately relevant for speaker identification. In fact, the problems associated with such an instrumental analysis are even more difficult if it is also considered that offender or suspects may not always be cooperative and also have reasons to disguise their vocal characteristics.

Exploring Speech Samples

A more straight forward use of speech recordings is to obtain direct linguistic information about the events under scrutiny. A first and obvious step is simply to listen to what is actually said and what may be meant by the speakers. This is not necessarily a simple task because such speech samples are often of low quality, both because the recording devices are typically of low quality and also because the acoustic environment can be quite adverse. In addition to the linguistic content of the speech samples per se, it is often important to establish the truthfulness of what is said. There is a wide range of situations where detecting deception may be of critical relevance for the actions to be taken or for the verdict to be issued. One possible scenario would be a kidnapper making a telephone call to blackmail or give instructions for ransom payment. For the sake of argument, let us assume that there are no other sources of information besides what is actually said during the short telephone call, i.e. that there are no clues as to what group membership or reasons that the kidnapper might have for the action. If the blackmailed victim was lucky enough to be able to record the conversation with the criminal, determining whether or not the abductor could be trusted with regard to some of its threats or promises might be important information. Another possible forensic scenario would be detecting deception in a witness' declarations or a suspect's statements but there are many other possible areas where the ability to detect deception by analyzing the speaker's voice would have obvious advantages. For instance, detecting deception for security purposes by voice analyses of the answers to screening questions, detecting false declarations in application calls for applications to different benefits or spotting deception in the voice of business partners are some of the areas where voice analyses capable of instrumental deception detection would have the prospect of wide and lucrative markets. Not surprisingly, such a prospect has indeed attracted the interest of some commercial actors who claim to have products that can detect lies (and a host of emotional states) through instrumental analyses of a speaker's voice. These products are marketed under several names but to all practical effects they can be "lie detectors".

"LIE DETECTION" BY VOICE ANALYSES

Two such "lie detectors" will be briefly described and discussed below – one that is claimed to rely on scientific research on physiological tremor (Lippold, 1971), a "voice stress analyzer" (VSA), and the other using a so-called

F. Lacerda

“Layered Voice Analysis” (LVA), patented by the selling company (Lieberman, 2003) – essentially reiterating the arguments presented already in 2007 by Eriksson and Lacerda (Eriksson and Lacerda, 2007).

Voice Stress Analyzers

According to website post by the National Institute for Truth Verification, Computerized Voice Stress Analysis (CVSA) is a crime-fighting technology that is successfully used to obtain confessions from suspects. Here is how the system is introduced:

Unlike the computer polygraph, the CVSA[®] requires no wires be attached to the subject being tested. The CVSA[®] uses only a microphone plugged into the computer to analyze the subject's responses. As the subject speaks, the computer displays each voice pattern, numbers it and saves each chart to file. Unlike the polygraph, drugs do not affect the results of the exam and there are no known counter-measures that will cause the ubiquitous “inconclusive” results associated with the polygraph. (www.cvsal.com/CVSA.htm; accessed 2013-01-22)

The analyzer is claimed to detect micro tremors occurring in association with lying:

Micro tremors are tiny frequency modulations in the human voice. When a test subject is lying, the automatic, or involuntary nervous system, causes an inaudible increase in the Micro tremor's frequency. The CVSA[®] detects, measures, and displays changes in the voice frequency. (www.cvsal.com/CVSA.htm; accessed 2013-01-22)

However, while Lippold's experimental studies provided evidence for micro tremors as part of the neurophysiological muscular control during static postures of the limbs (Lippold, 1970; Lippold, 1971), the claims made by the National Institute for Truth Verification to promote CVSA have never been supported by scientific experiments. There are many reasons to be skeptical about their claims. One of the reasons to be skeptical comes from the long chain of assumptions that would have to be met for CVSA to work. Indeed, assuming, for the sake of argument, that micro tremors in voice do exist, it would also be necessary to demonstrate that they really are affected by stress and finally that lying always is stressful. In addition, as far as micro tremors of the laryngeal muscles are concerned, experimental research specifically designed to measure micro tremors has failed to demonstrate their existence (Shipp and Izdebski, 1981). Given that none of these components of the logical chain has actually been proven, the claims made in the CVSA advertisement must be regarded as ungrounded – a notion that is emphasized by the outcomes of several controlled empirical studies, suggesting that the system performs at chance level, or below (see Kreiman and Sidtis, 2011, Table 10.2, p 373).

An important question to be asked is why the notion of lie-detection based on voice stress analysis continues to be used by criminal investigators although its lack of scientific bases and successful performance was pointed out already more than three decades ago (Horvath, 1982). One reason may be, as suggested in the website of the National Institute for Truth Verification, that CVSA simply helps criminal investigators to “get confessions”. If criminal investigators have no knowledge of the scientific problems associated with their VSA devices and if they already are convinced that the suspect is guilty, then that all that is missing is just a “confession” and it really does not matter if the “lie-detection” system works or not, as long as the suspect believes it does. However, there are good reasons to question the ethic aspects of such procedures. In addition to the obvious risk that the investigator runs of being trapped into a dangerous and ineffective prejudice, there is also the strong possibility of planting false memories in the suspect, which effectively impair the value of the very information provided. Another bizarre implication of the “getting confessions” approach is the assumption that the investigator already knows the answer, suggesting that the use of a lie-detection device should be superfluous. At last, but not the least, hiding behind an “objective” analysis device is understandably comfortable, especially if such a device is promoted and perceived as “cutting-edge” technology.

LVA Technology

The LVA technology was created and patented in 2003 by Amir Lieberman, the founder of Nemesysco, an Israeli company marketing voice analysis software for emotion detection and risk assessment. The following information is posted at the company's website:

Using voice analysis for emotion detection and risk assessment

Nemesysco is dedicated to developing advanced and non-invasive investigation and security tools, fraud prevention solutions, CRM applications, consumer products, and psychological diagnostic tools.

All Nemesysco's products and services are based on **Layered Voice Analysis (LVA)**, our proprietary and patent protected voice analysis technology. (www.nemesysco.com; accessed 2013-01-22)

F. Lacerda

Nemesysco sells a wide range of products for a number of purposes, including forensic investigations and declares that all their products and services are based on Layered Voice Analysis (LVA). For an evaluation of the scientific status of Nemesysco's products it is therefore crucial to investigate the working principles of LVA. Inspection of their LVA patent undoubtedly reveals that their creative analysis technology completely lacks plausibility (Eriksson and Lacerda, 2007). Indeed, while the VSA technology is based on a misinterpretation and unfounded exploration of scientific results that might have been plausible, Nemesysco's LVA technology pushes "voice analysis" to the limits of final absurdity by relying on *ad hoc* measures that simply cannot provide any relevant information (Lieberman, 2003). The LVA technology uses successive and independent triplets of 8-bit sound samples, digitized at 11.025 kHz (Lacerda, 2009; 2012). Obviously, there is no valid emotional information within the 272 μ s time window corresponding to the three samples picked up by LVA. Before the "analysis", the sound samples are "filtered" by an integer division by 3, yielding a final representation in 256/3 quantization steps which results into roughly a 6 bit samples after "filtering". The analysis consists in counting the number of "thorns" as well as a crude statistics on the number and length of "plateaus" formed by sequences of triplets where consecutive samples do not differ by more than arbitrary threshold. From this counting, the program issues a range of estimates that are supposed to be related to the speaker's emotional state.

An important question is why people tend to be convinced of these gadgets' validity in spite of the total absence of a plausible working principle. A recently concluded study at the Stockholm University's Phonetics Laboratory suggests that it is the belief on an underlying working principle that leads subjects into attributing meaning to certain types of random outcomes. In fact, LVA should be viewed as a "voice-controlled" random generator whose output has probably been tinkered to produce statistically acceptable results. Since LVA tends to issue "lie" outputs as the length of the so-plateaus increases, hesitations and lowering of the speaker's fundamental frequency will tend to result in statistically longer plateaus, which led the system to issue "lie" warnings. Thus, although LVA cannot carry out any proper voice analysis, the system's instability is likely to be a key factor into creating the false belief that LVA is actually working. Indeed, LVA's secret of success is the same as fortunetellers'. By offering a range of spurious outputs, the user is led into making sense of the noise by creating her/his own interpretation bias that allows regarding as positive evidence the "favorable" outputs while dismissing as noise outputs that do not fit in.

LVA's applicability in forensic domains is therefore similar to that of VSA. Although for different reasons, both systems lack scientific bases and can only produce irrelevant results. Nevertheless, in a context of criminal investigation where the goal is just "to get a confession", these system's spurious random outputs can indeed prompt a confession if the suspect is led into believing that the system actually works.

LIE-DETECTION USING VOICE ANALYSES: SCIENCE OR PSEUDOSCIENCE?

Commercializing these voice analysis products is a lucrative business. In principle, all that is needed are some programming skills and a professionally designed website. The validity of the working principle can be masked for the general public by presenting instead anecdotic "evidence", preferably from people who may be perceived as having some authority in the matter. Not surprising, both the National Institute for Truth Verification and Nemesysco post in their websites extensive documentation intended to support their VSA- and LVA-based products based on customers' testimonies. In the case of Nemesysco, the company has managed to involve finance researchers who have been publishing in high rank finance journals (Mayew and Venkatachalam, 2012). These finance researchers have engaged into voice analyses of conference calls on financial transactions to determine how emotional information conveyed by the participants' voices might be relevant for financial purposes. This might have been a legitimate and potentially interesting research goal if Mayew and Venkatachalam had used a scientific approach but they obviously lack competence in speech or voice analysis and unfortunately wasted their research efforts because they carried out their voice analyses using the LVA-technology (Lacerda, 2012).

The final question is whether or not voice stress analysis can be properly used for forensic purposes. In principle it could, provided the methodology is transparent and its working principles are scientific sound. This involves both the validity of the measurements performed on the speech signal and the demonstration that those measurements are indeed relevant for the purposes of, for instance, lie-detection. There is a long sequence of steps involved in this process and all of them have to be clearly supported by scientific evidence and tested by scientific methodology. Until then, if VSA and LVA are the only two available options, it is best not to use them at all and bet instead on intelligent and scientifically grounded education programs for criminal investigators.

F. Lacerda

ACKNOWLEDGMENTS

This work was supported by Stockholm University's research resources.

REFERENCES

- Eriksson, A., and Lacerda, F. (2007). "Charlatanry in forensic speech science: A problem to be taken seriously," *International Journal of Speech, Language and the Law* **14**, 169-193.
- Horvath, F. (1982). "Detecting deception: the promise and the reality of voice stress analysis," *Journal of Forensic Sciences* **27**, 340-351.
- Kreiman, J., and Sidtis, D. (2011). "Miscellany: Voice in Law Enforcement, Media and Singing," in *Foundations of Voice Studies* (Wiley-Blackwell), pp. 361-397.
- Lacerda, F. (2009). "LVA technology : The illusion of "lie detection", " in *Fonetik 2009* (Universitetservice US-AB, Stockholm), pp. 220-225.
- Lacerda, F. (2012). "Money Talks: The Power of Voice: A critical review of Mayew and Ventachalam's The Power of Voice: Managerial Affective States and Future Firm Performance," in *PERILUS* (Stockholm University, Stockholm), pp. 1-10.
- Lieberman, A. (2003). "Layered Voice Analysis (LVA)," (US patent).
- Lippold, O. (1971). "Physiological tremor," *Scientific American* **224**, 65-73.
- Lippold, O. C. (1970). "Oscillation in the stretch reflex arc and the origin of the rhythmical, 8-12 C-S component of physiological tremor," *J Physiol* **206**, 359-382.
- Mayew, W. J., and Venkatachalam, M. (2012). "The Power of Voice: Managerial Affective States and Future Firm Performance," *The Journal of Finance* **67**, 1-44.
- Popper, K. (1992). *The logic of scientific discovery* (Routledge, London).
- Shipp, T., and Izdebski, K. (1981). "Current evidence for the existence of laryngeal macrotremor and microtremor," *Journal of Forensic Sciences* **26**, 501-505.